

Klasifikasi Karakter Tulisan Tangan berdasarkan pola segmen

20 Nopember 2010

Nuryuliani, Lulu C Munggaran

Fakultas Ilmu Komputer dan Teknologi
Informasi Universitas Gunadarma
Universitas Gunadarma Depok, Indonesia
nryulia@staff.gunadarma.ac.id,
lulu@staff.gunadarma.ac.id

Ulva Choyrivanie

Fakultas Ilmu Komputer dan Teknologi
Informasi Universitas Gunadarma
Universitas Gunadarma Depok, Indonesia
amethyst_gerrard@yahoo.co.id

Abstract

Salah satu tahap dari proses pengenalan karakter tulisan tangan adalah klasifikasi karakter tulisan tangan yang diawali dengan pengambilan data, segmentasi, klasifikasi sampai pencocokan. Fokus pada penelitian ini adalah tahap klasifikasi. Proses klasifikasi karakter tulisan tangan pada penelitian ini menggunakan metode ekstraksi ciri berbasis pola segmen. Untuk mempermudah proses klasifikasi, dilakukan proses filterisasi dan segmentasi terhadap karakter input. Klasifikasi karakter dikelompokkan menjadi 7 cluster berdasarkan pola segmen tiap karakter. Pada hasil segmen dari tiap karakter dilakukan pengkodean pola sehingga bisa dicocokkan dengan data yang ada dalam database. Hasil penelitian ini mampu melakukan klasifikasi dari 26 karakter acuan dengan keakuratan yang tinggi.

Keywords: Karakter; Klasifikasi; Pola Segmen; Tulisan Tangan.

1 Pendahuluan

Teknik pengenalan tulisan tangan dapat dilakukan secara online dan offline. Teknik yang pertama, dikenal sebagai pengenalan secara offline. Pada teknik ini hasil yang didapat dari hasil pemindaian berupa gambar tanpa adanya informasi sementara. Untuk dapat dilakukan proses pengenalan terhadap karakter tulisan yang dibuat dibutuhkan preprocessing seperti ekstraksi kontur, penipisan dan skeletonizing. Proses ini dapat menyebabkan berkurangnya akurasi terhadap pengenalan karakter.

Teknik pengenalan tulisan yang kedua seringkali dikenal sebagai pengenalan online. Pada pengenalan secara online, pengguna secara fisik dihubungkan dengan komputer melalui tetikus, pena elektronik, alat sensitif sentuhan atau tablet digitizer, kemudian tulisan tangan tersebut direkam sebagai proses yang tergantung dari waktu (time-dependent process). Data yang diambil secara online mempunyai kelebihan yaitu terdapatnya informasi sementara atau dinamis dari tulisan yang dibuat. Informasi ini terdiri dari jumlah gerakan (stroke), urutan getaran, arah tulisan untuk setiap gerakan dan kecepatan tangan setiap gerakan. Dengan menggunakan informasi sementara dapat meningkatkan keakuratan dalam pen-

genalan karakter.

Tappert, C. C., mengungkapkan bahwa terdapat beberapa metode yang digunakan untuk pengenalan bentuk karakter tulisan secara online diantaranya adalah analisis ciri, urutan waktu dari zona atau arah, pencocokan kurva, pengkodean gerakan, analisis by sintesis, perbedaan pasangan dan lain sebagainya[1]. Untuk dapat melakukan pengenalan karakter tulisan tangan, salah satu tahap dari proses adalah klasifikasi karakter tulisan tangan yang diawali dengan pengambilan data, segmentasi, klasifikasi sampai pencocokan. Fokus pada penelitian ini adalah tahap klasifikasi. Proses klasifikasi karakter tulisan tangan pada penelitian ini menggunakan metode analisa ciri berbasis pola segmen

2 Tinjauan Pustaka

2.1 Segmentasi Tulisan Tangan

Plamondon telah mengusulkan teori kinematik yang dapat digunakan sebagai kerangka umum segmentasi untuk mempelajari dan menganalisa tulisan tangan [7]. Teori ini didasarkan pada persamaan delta-Lognormal yang menguraikan kecepatan suatu gerak yang selanjutnya dapat digu-

nakan untuk menjelaskan karakteristik dasar dari sebuah stroke. Stroke adalah gerakan menulis mulai dari pena di letakkan sampai pena diangkat dari permukaan tablet [1]. Menurut teori Plamondon, tulisan tangan terdiri dari kurva linear dan stroke sudut yang terbentuk saat terjadinya pergerakan cepat pada manusia. Salah satu cara paling efisien untuk mengekstrak stroke adalah melakukan percobaan analysis-by-synthesis atas komponen jejak gerakan pena digital yang dihasilkan selama gerakan menulis.

Stroke atau segmen merupakan hasil dari proses segmentasi yang membagi keseluruhan input tulisan tangan menjadi potongan yang lebih kecil. Input tulisan tangan dapat berupa bentuk teks, kata, atau karakter. Segmentasi mempunyai tujuan membagi karakter tulisan tangan kedalam beberapa bagian untuk mendapatkan ciri dari karakter tersebut. Ciri ini selanjutnya akan digunakan untuk proses pengenalan karakter tulisan tangan. Jika proses segmentasi tidak dapat menentukan secara tepat posisi titik potongan segmen, maka sistem aplikasi akan sulit atau kurang tepat dalam mengenali tulisan tangan.

Langkah awal dari metode segmentasi tulisan tangan adalah melakukan preprocessing, yaitu smoothing yang berguna untuk menghilangkan noise yang ada pada tulisan tangan. Smoothing dapat dilakukan dengan menggunakan wavelet ataupun filter lainnya. Karakteristik yang biasa digunakan sebagai dasar untuk melakukan segmentasi adalah koordinat x dan y [6], kecepatan linier dan kecepatan sudut serta sudut pergerakan pena [2].

Segmentasi tulisan tangan ditujukan untuk lebih mempermudah proses pengenalan tulisan tangan. Dalam penelitian ini, menggunakan segmentasi yang dikembangkan oleh Suryarini [8]. Dari 26 karakter tulisan tangan huruf latin terdapat 25 jenis segmen acuan. Contoh segmentasi standar pada beberapa karakter yang dilakukan secara manual berdasarkan pendapat sejumlah peneliti [5],[9] dapat dilihat pada gambar 1.

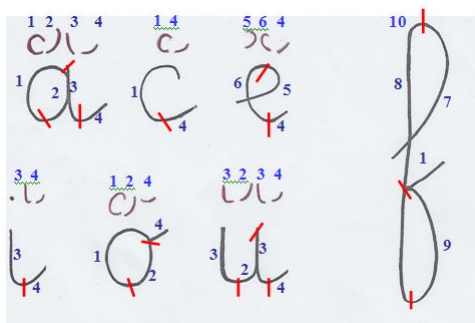
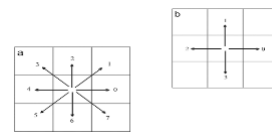


Figure 1: Contoh segmentasi pada karakter a,c,e,f,i,o dan u [5],[9]

2.2 Penentuan Kode Freeman

Penentuan Kode Freeman dilakukan setelah diperoleh batas tepi objek dari citra atau sinyal dan digunakan untuk menggambarkan bentuk luar dari obyek dengan serangkaian kode dari garis-garis pada batas tepi obyek berdasarkan panjang dan arah tertentu [9].

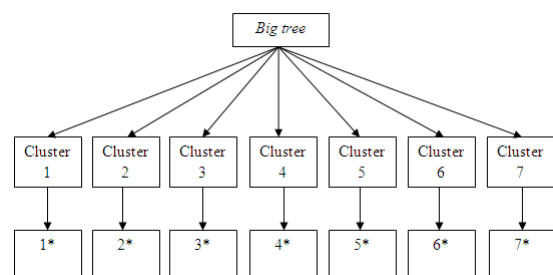
Arah dari representasi batas tepi obyek menggunakan skema 4-arah mata angin atau delapan arah mata angin yang telah diberi angka tertentu, seperti yang digambarkan pada skema arah mata angin pada gambar 2.



Gambar 2: (a). Skema 8 arah mata angin, (b). Skema 4 arah mata angin

2.3 Klasifikasi Karakter

Untuk mempermudah dan mempercepat akses serta proses pencocokan maka perlu dilakukan klasifikasi karakter berdasarkan pada pola segmen. Hal ini dilakukan karena terdapat beberapa segmen dalam sejumlah karakter yang memiliki pola yang sama. Klasifikasi dilakukan dengan cara mengelompokkan semua karakter yang memiliki kesamaan ciri dan pola mulai pada segmen pertama. Selanjutnya, kelompok karakter ini dipecah menjadi sub-kelas berdasarkan kesamaan ciri dan pola segmen kedua. Hal yang sama dilakukan untuk segmen ketiga dan demikian seterusnya. Dengan demikian akan terbentuk kelas dan sub-kelas dalam urutan pohon hirarki seperti yang ditunjukkan oleh gambar 3.



Gambar 3: Pohon clustering karakter berdasarkan bentuk segmen

3 Metodologi Penelitian

3.1 Penentuan Obyek Penelitian

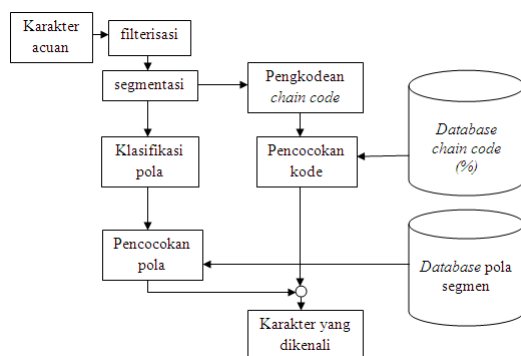
Data yang akan dijadikan sebagai obyek penelitian di peroleh dari peralatan masukan berupa tablet digitizer. Tulisan yang diambil mengikuti format yang telah ditentukan baik bentuk huruf dan arah menulis. Karakter yang digunakan adalah huruf a sampai dengan huruf z. Data yang dihasilkan dari tablet digitizer berupa koordinat X, koordinat Y dan tekanan.

3.2 Bahan dan/atau alat Penelitian

Bahan dan/ atau alat yang digunakan dalam penelitian ini adalah: Tablet digitizer yang dilengkapi dengan pena elektronik, Perangkat lunak MATLAB Image Processing Toolkits versi 7 dan seperangkat komputer.

3.3 Metode Eksperimen

Secara umum, skema dari metode klasifikasi karakter tulisan tangan terbagi atas dua proses, yakni proses pembentukan basis data karakter acuan dan proses klasifikasi. Pada dasarnya program klasifikasi karakter tulisan tangan merupakan bagian dari sistem pengenalan tulisan tangan yang diawali dengan proses pengambilan data, segmentasi, klasifikasi sampai pengenalan[4]. Skema proses pengenalan karakter tulisan tangan dapat dilihat pada Gambar 4.



Gambar 4: Skema proses pengenalan karakter[4]

Langkah-langkah yang dilakukan adalah sebagai berikut:

3.4 Pengkodean Pola Segmen

Karakter yang sudah di segmentasi kemudain diko-dekan kedalam chain code dan pola segmen agar dapat dicocokkan dan memeriksa kesesuaian dalam database. Database chain code berisi probabilitas kemunculan tiap segmen dalam setiap kode chain

code. Untuk lebih jelasnya dapat dilihat pada gambar 5 .

Karakter	Kode 1	Kode 2	Kode 3	Kode 4	Kode 5	Kode 6	Kode 7	Kode 8
A1	0.0312	0.0000	0.0057	0.0439	0.2424	0.2395	0.3494	0.0879
C1	0.0893	0.0000	0.0037	0.0384	0.2295	0.2723	0.2413	0.1255
D1	0.0138	0.0000	0.0061	0.0697	0.2714	0.2482	0.3081	0.0648
E2	0.0952	0.0000	0.0000	0.0000	0.0440	0.2686	0.4762	0.1160
G1	0.0433	0.0000	0.0119	0.0422	0.2771	0.2489	0.2987	0.0779
O1	0.1040	0.0000	0.0000	0.0000	0.1312	0.3131	0.3094	0.1423
Q1	0.0488	0.0000	0.0137	0.0615	0.2012	0.3291	0.2549	0.0908

Gambar 5: Daftar probabilitas dari segmen 1

Keterangan:

- : nilai batas bawah
- : nilai batas atas

3.5 Pengklasifikasian Pola

Pola segmen dari karakter diklasifikasikan menjadi 7 cluster berdasarkan pola segmen pertama dari setiap karakter, kemudian seterusnya sampai huruf dikenali. Data untuk jenis segmen dapat dilihat pada gambar 6 .

Karakter	Jumlah segmen	Jenis segmen	Karakter	Jumlah segmen	Jenis segmen
Huruf A	4	1,8,5,9	Huruf N	5	3,25,18,19,9
Huruf B	4	2,14,8,15	Huruf O	3	1,26,15
Huruf C	2	1,9	Huruf P	5	4,12,22,19,9
Huruf D	4	1,8,10,9	Huruf Q	3	1,8,12
Huruf E	3	7,1,9	Huruf R	4	4,15,5,9
Huruf F	4	2,14,16,15	Huruf S	2	4,23
Huruf G	4	1,8,11,13	Huruf T	2	5,9,15
Huruf H	5	2,17,18,19,9	Huruf U	4	5,8,5,9
Huruf I	2	5,9	Huruf V	4	3,19,8,15
Huruf J	3	4,11,13	Huruf W	6	3,19,8,5,8,15
Huruf K	6	2,17,18,20,21,9	Huruf X	3	6,8,9
Huruf L	3	2,14,9	Huruf Y	5	3,19,8,11,13
Huruf M	7	3,25,18,25,18,19,9	Huruf Z	5	4,15,4,23,13

Gambar 6: Jumlah segmen dari 26 huruf alfabet

3.6 Pencocokan

Pada sub-bab ini akan dijelaskan proses pencocokan kode segmen dari pola segmen setiap karakter dalam pengklasifikasian karakter. Pada penelitian ini menggunakan seleksi pola pada segmen pertama pertama tiap karakter. Segmen pertama tiap karakter diseleksi untuk menentukan kluster dari karakter. Setelah ditentukan klusternya, nilai tersebut digunakan untuk melakukan seleksi segmen berikutnya sehingga huruf input dapat diklasifikasi dan dikenali.

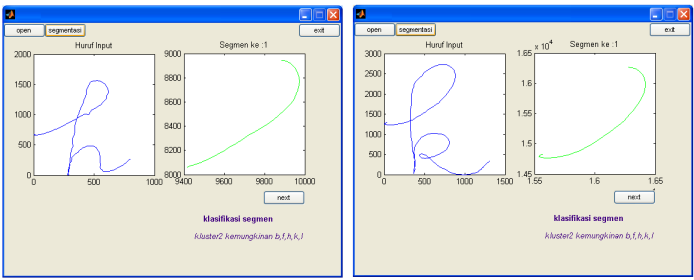
4 Hasil Eksperimen

Hasil dari penelitian berdasarkan uji coba pada pengambilan data dari setiap karakter sudah mendapatkan hasil yang diharapkan. Setelah menyelesaikan perencanaan dan perancangan software, selanjutnya untuk mengetahui serta meyakinkan bahwa seluruh sistem perancangan perangkat telah be-

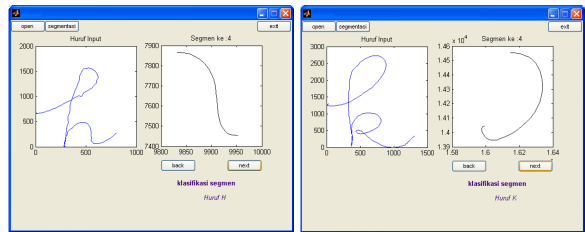
kerja dengan baik dilakukan pengujian pada software tersebut.

Pengujian ini bertujuan untuk membuktikan apakah program aplikasi ini mampu mengklasifikasi setiap karakter yang dimasukkan. Pada bagian ini akan dijelaskan mengenai proses klasifikasi karakter tulisan tangan.

Tampilan program setelah diruning dapat dilihat pada gambar 7 dan 8.



Gambar 7: Tampilan awal dari huruf input h dan k



Gambar 8: Tampilan hasil klasifikasi dari huruf h dan k

Dari hasil uji coba diatas, aplikasi berhasil melakukan klasifikasi dengan baik. Uji coba juga dilakukan pada karakter yang lain, keakuratan dari program dapat dilihat pada tabel 3

Setelah dilakukan uji coba pada 26 jenis huruf alfabet dapat diambil kesimpulan dari keakuratan program sudah cukup tinggi dengan dapat dikenalnya semua karakter huruf alfabet.

5 Kesimpulan dan Saran

5.1 Kesimpulan

Klasifikasi karakter tulisan tangan pada penelitian ini menggunakan metode ekstraksi ciri dari pola segmen. Karakter input difilter terlebih dahulu sebelum dilakukan segmentasi. Kemudian karakter dikelompokkan menjadi 7 cluster berdasarkan segmen pertama tiap karakter. Pada hasil segmen dari tiap karakter dilakukan pengkodean pola segmennya sehingga bisa dicocokkan dengan data yang ada dalam database. Aplikasi ini mampu melakukan klasifikasi dari 26 karakter acuan dengan keakuratan yang tinggi.

Tabel 1: Tabel kecocokan percobaan 26 jenis huruf alfabet

Karakter Input	pencocokan
Huruf a	cocok
Huruf b	cocok
Huruf c	cocok
Huruf d	cocok
Huruf e	cocok
Huruf g	cocok
Huruf h	cocok
Huruf i	cocok
Huruf j	cocok
Huruf k	cocok
Huruf l	cocok
Huruf m	cocok
Huruf n	cocok
Huruf o	cocok
Huruf p	cocok
Huruf q	cocok
Huruf r	cocok
Huruf s	cocok
Huruf t	cocok
Huruf u	cocok
Huruf v	cocok
Huruf w	cocok
Huruf x	cocok
Huruf y	cocok
Huruf z	cocok

Klasifikasi karakter tulisan tangan merupakan bagian dari sistem pengenalan karakter tulisan tangan yang diawali tahap pengambilan data, segmentasi, klasifikasi sampai pencocokan. Sehingga, penelitian ini kedepannya dapat dijadikan referensi dalam sistem pengenalan tulisan tangan khususnya dalam tahap pencocokan.

5.2 Saran

Untuk penelitian di masa kedepan diharapkan proses klasifikasi ini dapat dilanjutkan ke tahap pengenalan tulisan tangan sehingga dapat dimanfaatkan untuk membantu proses pembelajaran. Dan untuk meningkatkan tingkat akurasi pengenalan, sebaiknya lebih memperkaya database dengan jumlah referensi yang lebih banyak.

References

- [1] C.Y. Sun dan Y. Wakahara C. C. Tappert. The state of the art in on-line handwriting recognition. In *IEEE Transaction on Pattern and Machine Intelligence.*, volume 17, August 1990.
- [2] M. Sacher-Reyes Mas. P. Sacher Garcia E. Gomez. Sanchez, Y.A. Dimitriadis. On-line

character analysis and recognition with fuzzy neural network. 1998.

- [3] Herbert Freeman. Computer processing of line-drawing images. In [1] *Computing Survey*, volume 6. March 1974.
- [4] Sarifuddin Madenda dan Michel Paindavoine Nuryuliani, Lulu C. Munggaran. Pendekatan kode rantai sebagai dasar pengenalan karakter. In *SNATI 2009*, UII-Universitas Islam Indonesia, 2009.
- [5] Michel Paindavoine. Discussion laboratoire le2i, universit   de bourgogne. Dijon-France, 2008.
- [6] R Plamondon. A model-based segmentation framework for computer processing of handwriting. In *IEEE Proceeding of ICPR*, 1996.
- [7] Moussa R Plamondon dan Djioua. Handwriting stroke trajectory variability in the context of the kinematic theory. advances in graphonomics. In 2005, editor, *Proceeding of IGS*, 2005.
- [8] Widodo Suryarini. Metode segmentasi karakter tulisan tangan on-line menggunakan karakteristik perubahan nilai koordinat y. Disertasi. Universitas Gunadarma, 2009.
- [9] Annie Vinter. Discussion laboratoire le2i, universit   de bourgogne. Dijon-France, 2008.